PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

# Building Knowledge-Based Systems for Detecting Man-Made Structures from Remotely Sensed Imagery

D. M. McKeown

*Phil. Trans. R. Soc. Lond. A* 1988 **324**, 423-435
doi: 10.1098/rsta.1988.0030

*Phil. Trans. R. Soc. Lond.* A **324**, 423–435 (1988)
*Printed in Great Britain*



# Building knowledge-based systems for detecting man-made structures from remotely sensed imagery

By D. M. McKeown, Jr

*Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania* 15213, *U.S.A.*

[Plates 1–3]

In this paper we review and discuss several emerging themes in the area of image-interpretation for automated cartography and digital mapping. Our primary observations are that the detection, interpretation, and analysis of man-made structures by remotely sensed imagery requires the integration of spatial knowledge with image-analysis and interpretation techniques. General knowledge of structural and spatial layouts for man-made structures, as well as *a priori* map knowledge, can be expected to constrain search during image analysis. Spatial databases consisting of geographically referenced facts should form the basis of *a priori* knowledge necessary to perform interpretation and have a dual role as the repository of partially compiled facts acquired during scene-analysis.

## 1. Introduction

Knowledge-based interpretation of remotely sensed data requires knowledge about the scene under consideration. Knowledge about the type of scene (airport, suburban housing development, urban city) can aid in low-level and intermediate-level image-analysis, and can be expected to drive high-level interpretation by constraining search for plausible consistent scene models. Loosely speaking, this knowledge base should contain known facts and spatial relations between objects in an area of interest, access to historical or *a priori* map knowledge, and methods to relate Earth coordinates to pixel locations in digital imagery.

In this paper we describe three experiments in knowledge-based interpretation. Each experiment uses spatial and structural knowledge at an increasingly higher level of detail. These experiments can be characterized as narrowly defined, vertically integrated image-analysis systems used to extract specific features such as roads, buildings and complex sites such as airports and urban environments. These programs are not 'general vision' systems because they capitalize on the task-specific nature of the problem domain to constrain the complexity of the world. Nevertheless, they are important datapoints in a domain where many important questions concerning knowledge representation and use largely remain unanswered.

The primary role for knowledge-based systems in the analysis of remotely sensed imagery is to provide constraints so that image analysis and interpretation tools can be used in spite of their inherently error-prone performance. The goal then, is to integrate rule-based systems with image-analysis techniques to constrain the search space of possible scene interpretations by using domain and general knowledge. These constraints can be characterized as 'what to look for' and 'where to look for it'. Currently, knowledge-based systems are most powerful (and successful) in narrow, well-defined task areas. It is certainly the case that to solve the general remote-sensing problem an analysis system will require use of general problem-solving capabilities and vast amounts of domain and common-sense knowledge, currently far beyond the capability of any research system. However, even without general problem-solving capabilities, there is much to be gained by the development of 'existence proof' knowledge-based interpretation systems. Current image-processing systems are incapable of high-level description of the results of image interpretation. The combination of a map-based world model and knowledge-based systems which have site-specific or task-specific knowledge can be used to bridge the gap between users and current state-of-the-art image-interpretation systems. The long-term goal of our research is to develop systems that can interact with a human cartographer or analyst at a highly symbolic level, maintain a database of previous events, and use expert-level knowledge to predict areas for fruitful analysis, and integrate the results of the analysis into a coherent model.

In the remainder of this paper we describe three experimental analysis systems that use low-level, intermediate-level and high-level spatial knowledge, respectively, to extract and interpret man-made structures from aerial imagery. These systems are MACHINESEG, a system for map-guided image segmentation, ARF, a system for finding and tracking roads, and SPAM, a rule-based system for airport scene interpretation. Each is linked by the common use of a spatial database system, MAPS/CONCEPTMAP, to provide access to and information concerning image, map, and terrain data. In the following section we briefly describe the MAPS database representation for spatial knowledge.

## 2. An overview of MAPS

The MAPS spatial database (McKeown 1983, 1984*a*, *b*, 1987) was developed between 1980 and 1984, supported by the DARPA Image Understanding Program as research into large-scale spatial databases and spatial knowledge representation. It is interesting that this system has expanded from its original research goal of developing an interactive database for answering spatial queries into a component of several knowledge-based image-understanding systems under development at Carnegie-Mellon University. MAPS is a large-scale image–map database system for the Washington, D.C. area that contains approximately 200 high-resolution aerial images, a digital terrain database and a variety of map databases from the United States Geological Survey (USGS) and the Defense Mapping Agency (DMA). MAPS has been used as a component for an automated road finder–follower, a stereo verification module, a 3-D scene-generation system, and a knowledge-based system for interpreting airport scenes in aerial images. In addition, MAPS has an interactive user-query component that allows users to perform spatial queries with high-resolution display of aerial imagery as a method for indexing into the spatial database. This capability to relate, over time, imagery at a variety of spatial resolutions to a spatial database forms a basis for a large variety of interpretation and analysis tasks such as change detection, map update and model-based interpretation.

Figure 1 shows the system organization of MAPS. Four databases are maintained within MAPS: a digital terrain database, a map database, a landmark database and an image database. A fifth database, CONCEPTMAP, consists of a schema-based representation for spatial entities and a set of procedural methods that provide a uniform interface to each of the four component databases for interactive users or application programs. It is this interface that allows us to represent and access image, map, terrain and collateral data in a manner that best suits the intrinsic structure of the data. At the same time the CONCEPTMAP databases provides uniform access to a variety of spatial data independent of the particular internal structure. This is in





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

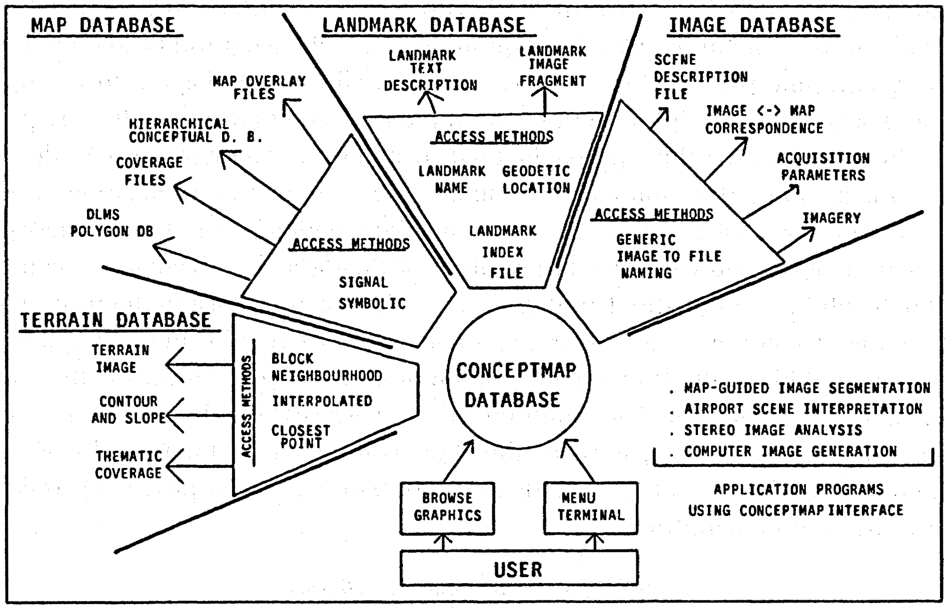


FIGURE 1. MAPS: system overview.

sharp contrast to methods proposed for uniform representation of image and cultural data such as raster data sets and regular decompositions such as quadtrees or k–d trees.

### 2.1. *Schema-based representation for spatial entities*

The CONCEPTMAP database uses a schema-based representation for spatial entities. The use of schemas (or frames) is a well-understood AI methodology for representing knowledge. Such a representation can be combined within several problem-solving methods such as semantic networks, scripts or production systems to construct a problem-solving system (Barr & Feigenbaum 1981). Each entity in the CONCEPTMAP database is represented by one concept schema and at least one role schema. A concept can represent any spatial object and associates a name with a set of attributes stored in the concept and role schemata. A concept such as 'washington d.c.' might have multiple role schemata defined, each representing a different view of the same spatial area, such as 'political', and 'demographic'. A concept such as 'georgetown university' might have multiple role schemata defined for each of the campus buildings or areas.

There are three unique identifiers generated by the CONCEPTMAP system which allow for indirect access to additional factual properties of concept or role schemata.

(i) The concept-id is unique across all concepts in all CONCEPTMAP databases. That is, given a concept-id one can uniquely determine the name of the spatial entity.

(ii) The role-id uniquely determines a role schema across all CONCEPTMAP databases.

(iii) The role-geographics-id uniquely determines a collection of points, lines or polygons in vector notation. Each point is represented as $<$ latitude, longitude, elevation $>$.

These identifiers are also used to index into other components of the MAPS database. For example, the concept-id is used to search for landmark descriptions of measured ground control points used during the calculation of transform functions for image-to-map and map-to-image correspondence. The role-id is used as the basic entity when building fast-access methods to the spatial data by using a hierarchy tree decomposition. The role-geographics-id is used to acquire the unique geographic position for a role schema as well as for linkage into the MAPS image database and segmentation files generated by human interaction or machine segmentation. There are three reasons for this approach. First, it allows CONCEPTMAP to handle very large databases with a minimal amount of information resident in the application process. The identifiers provide a level of indirection to the actual data, which is stored in a variety of formats and may or may not be present for a large subset of the database. Second, we can achieve a great deal of flexibility and modularity in processes which communicate about spatial entities. Given the name of a CONCEPTMAP database, a concept-id or role-id uniquely determines the entity in question. This facilitates the construction of application programs with simple query structures, requiring a minimum of communication overhead. Finally, given this decoupling from the CONCEPTMAP database, each of the MAPS component databases – image database, terrain database, landmark database and map database – may be physically resident on a different workstation or mainframe.

### 2.2. *A geodetic frame of reference*

An implicit requirement crucial to successful application of spatial knowledge for image analysis is that the metrics used by the analysis system be defined in cartographic coordinates, such as < latitude, longitude, elevation > , rather than in an image-based coordinate system. Systems that rely on descriptions such as 'the runway has area 12000 pixels' or 'suburban homes are between 212 and 345 pixels in area' are useless except for (perhaps) the analysis of one image. Further, spatial analysis based on the semantics of above, below, left-of, right-of, etc., are also inappropriate for general interpretation systems. To apply metric knowledge, one must relate the world model to the image under analysis. This should be done through map-to-image correspondence with camera models. We can directly measure ground distances, areas and absolute compass direction, and recover crude estimates of height using a camera model computed for each image under analysis. The MAPS system allows us to express spatial knowledge in ground-bases metrics, which in turn allows us to work with imagery and map data at a variety of scales and resolutions.

In the following sections we describe the first of our three experimental analysis systems, MACHINESEG. It can be characterized as using low-level spatial knowledge, primarily two-dimensional shape information, coupled with general positional information to guide region analysis and segmentation.

## 3. Map-guided image segmentation

MACHINESEG (McKeown & Denlinger 1984) is a program that performs map-guided image segmentation. It uses map knowledge to control and guide the extraction of man-made and natural features from aerial imagery using a region-growing image-analysis technique. MACHINESEG uses the CONCEPTMAP database from the MAPS system as its source of map knowledge. Figure 2 shows the interaction between the map database and image-processing and feature-extraction tools. Map knowledge can be used to represent generic shapes, sizes and spectral properties typical to a large class of objects such as roads, or can describe specific features such as known buildings, where geodetic position may be known as well as the more

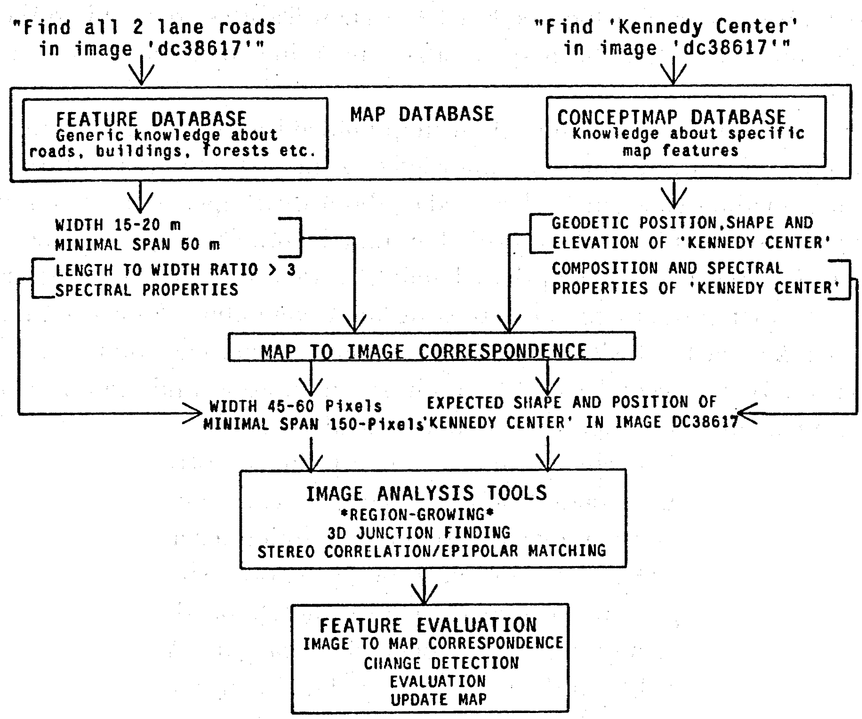


FIGURE 2. MACHINESEG: map-guided feature extraction.

general structural properties. For the latter we have been able to segment a diverse set of cultural features such as buildings, reservoirs and roads, using map-to-image correspondence to project map-based descriptions on to a new image under analysis. This projection generally only provides a coarse idea of the actual position of the feature, but greatly constrains search in the image. MACHINESEG uses the following components of the MAPS system.

(i) CONCEPTMAP to retrieve shape and position models.

(ii) Map-to-image correspondence to project models on to new imagery and image-to-map correspondence to calculate metric distances and areas.

(iii) CONCEPTMAP to store extracted features from several images before interactive editing and integration into the database.

It is important to characterize what we mean by 'map-guided' image segmentation. Map-guided image segmentation is the application of task-independent spatial knowledge to the analysis of a particular image, using an explicit map-to-image correspondence derived from camera and terrain models. Map-guided segmentation is not interactive editing or computation of descriptions in the image domain, because these descriptions are valid for only one specific image. As we have described, each role schema contains a role-geographic-id which points to a geodetic description ( < latitude, longitude, elevation > ) for each map entity in the CONCEPTMAP database. This description is in terms of points, lines and polygons, or collections of these primitives. Features such as buildings, bridges and roads have additional attributes describing their elevation above the local terrain, as well as their composition and appearance. The location of each map feature in the database can be projected on to a new image by a map-to-image correspondence maintained by MAPS. Likewise, a new map feature can be projected on to the existing image database. If camera-model errors are known, one can directly calculate an uncertainty for image-search windows. Further, as new features are acquired their positions can be directly integrated into the map database by using image-to-map correspondence procedures.

Figure 2 shows a schematic description of the map-guided feature-extraction process in MAPS. There are two methods for applying map knowledge to the extraction of features from aerial imagery. The first method uses generic knowledge about the shape, composition and spectral properties of man-made and natural features. This may be provided by the knowledge base of the application. The second uses map-based template descriptions. These descriptions are stored in the CONCEPTMAP database and represent knowledge about known buildings, roads, bridges, etc. This knowledge includes geodetic position, shape, elevation, composition and spectral properties. In the second case, the position, orientation and scale are constrained, whereas in the first, only the scale can be determined. For both to operationalize spatial knowledge for the analysis of a particular image, a map-to-image correspondence is performed.

Photographs shown in figure 3, plate 1, show the sequence of region-merging steps performed automatically by MACHINESEG using a generic shape description for a set of buildings near the Ellipse park in Washington, D.C. Photograph 0 shows the original image. Photographs 1–5 show how the region merges are evaluated against the shape description, those regions satisfying the shape criteria being marked as solid filled polygons. While a large number of potential regions are generated from the initial image (1), only a small number correspond to buildings. Photograph 5 shows the final result of all seven buildings correctly segmented, with two incorrect segmentation regions generated.

It should be stated that MACHINESEG does not have a model of buildings, or the relation of buildings to roads, or any particular high-level description of the image under analysis. Such models or expectations must be provided by other processes, such as an interactive user, or an analysis system. What makes MACHINESEG a powerful image-analysis tool is that it can use shape descriptions and search among the region segmentation for those regions that best match the segmentation in terms of shape, size and orientation properties. Thus, MACHINESEG is a good example of a low-level generator of plausible regions that must be analysed by intermediate and high-level processes within the context of a particular task. It can be argued that more specific knowledge about the structure of buildings and roads should be brought to bear at the low level. Our philosophy has been to avoid that feature-specific approach during the segmentation phase and to defer such analysis until a coarse overview of the entire scene can be accomplished.

In the following section we describe a system, ARF, that produces intermediate-level and high-level descriptions of roads in high-resolution aerial imagery. It uses knowledge about the structure of roads and employs multiple methods to improve overall system performance. In contrast to MACHINESEG, which can also be used to segment linear features, such as roads, railroad lines and certain drainage patterns, ARF relies on a very specific model of roads to achieve a much higher reliability and performance levels than could be achieved with a general region-based system such as MACHINESEG.

## 4. Cooperative methods for road tracking

Road finding in remotely sensed imagery has often been equated with linear feature extraction. The rationale was that finding linear features in imagery either by region extraction or line finding was equivalent to finding roads. Further, line finding also worked for other types

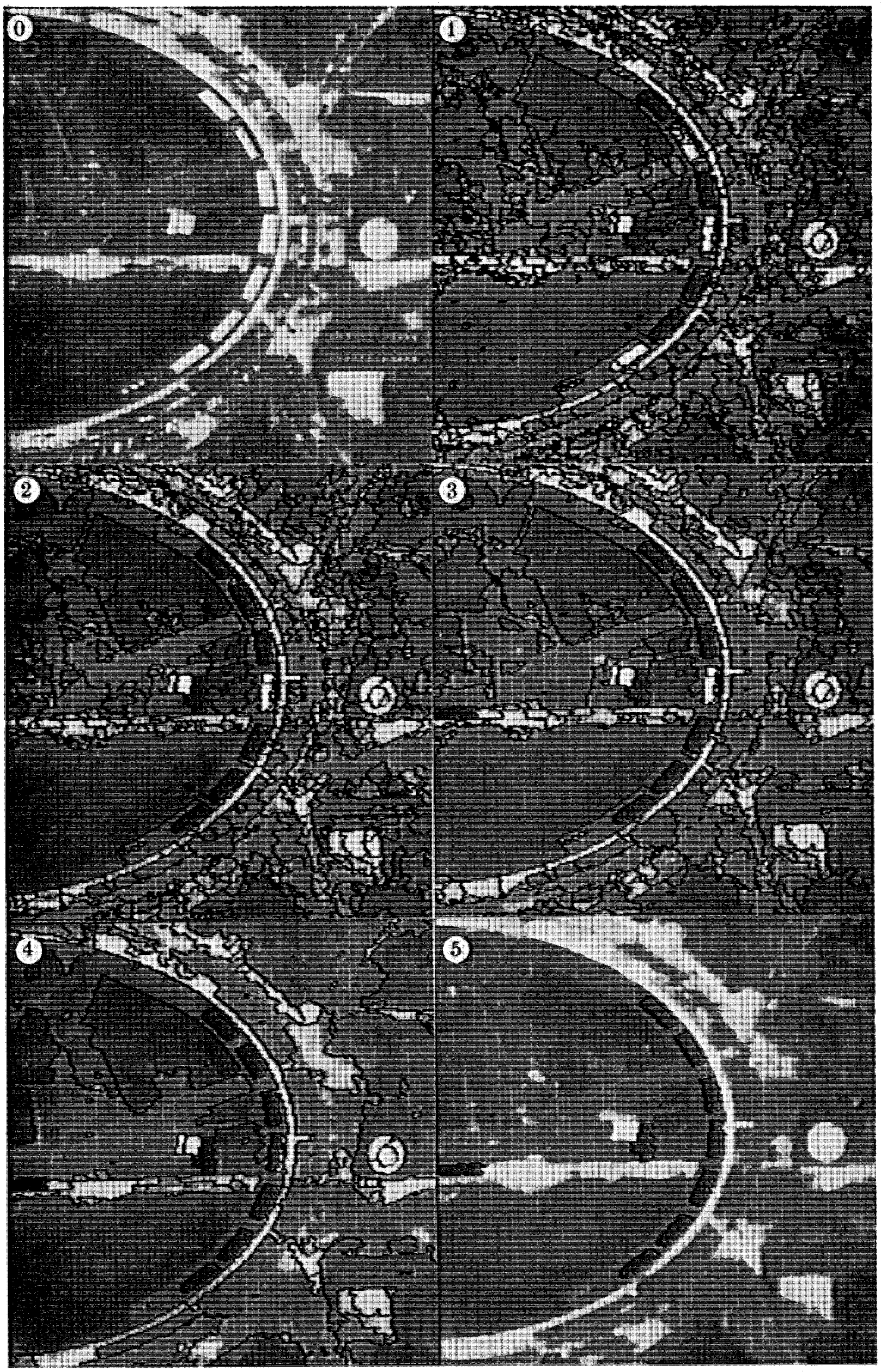


FIGURE 3. Sequence of image segmentation for small buildings.

(*Facing p.* 428)

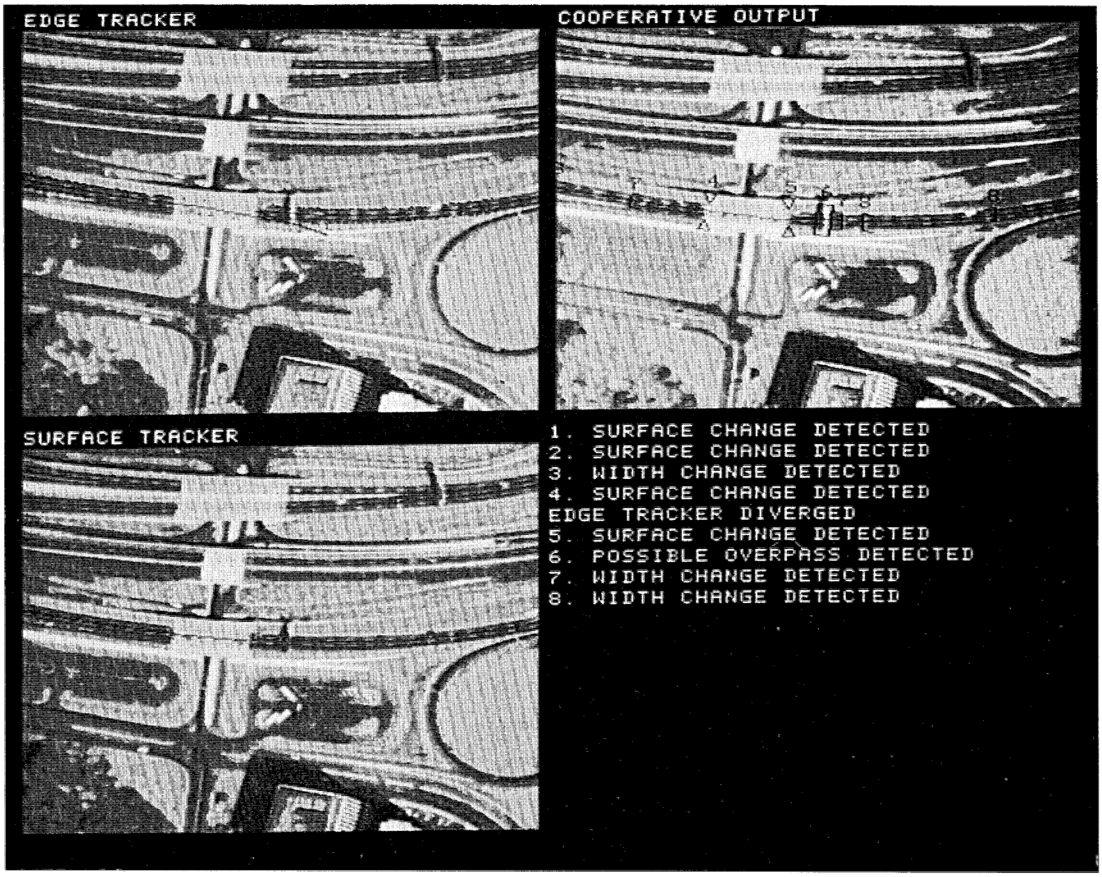


FIGURE 5. ARF: edge tracker fails on low-contrast overpass, successfully restarted from surface tracker.

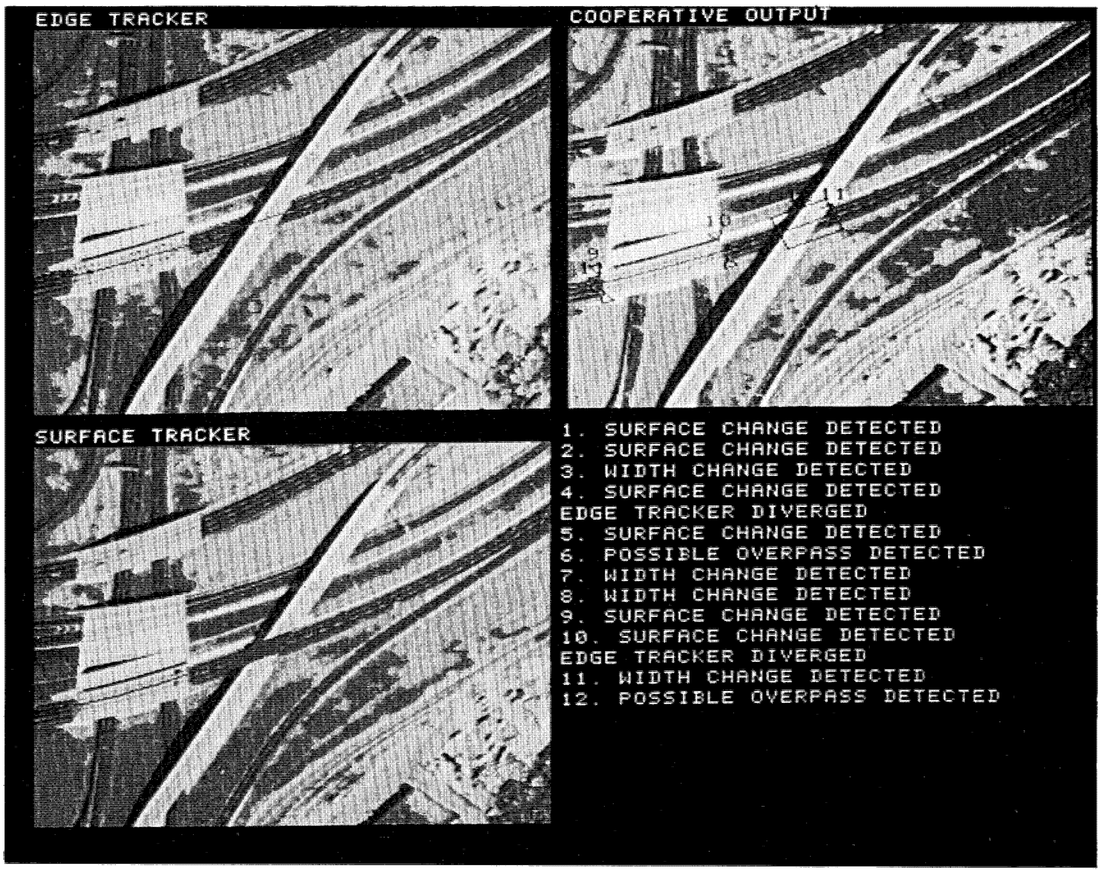


FIGURE 6. ARF: successful detection of overpass and surface material change due to bridge deck.

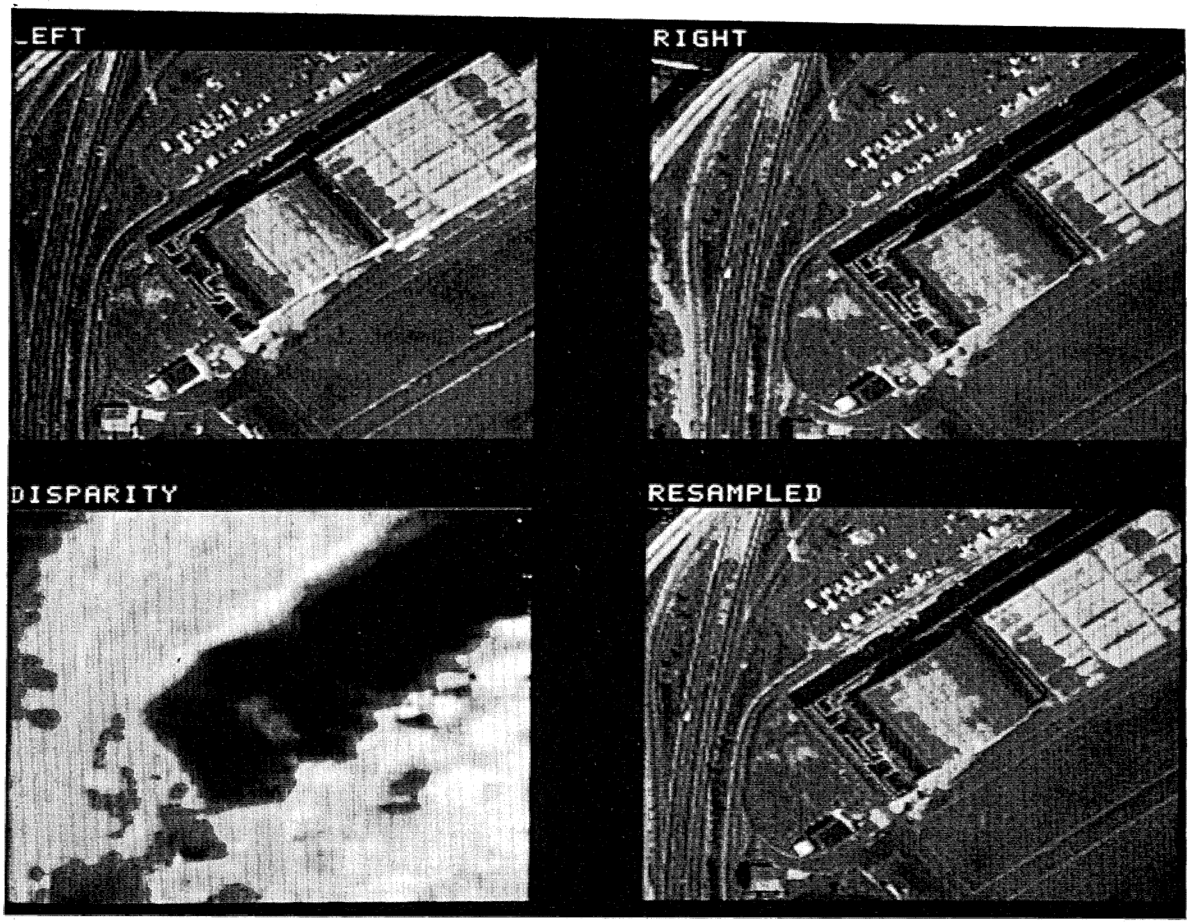


FIGURE 10. Stereo verification in airport scene analysis.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

of linear features such as drainage, bridges and railroads. This view was appropriate considering the relatively low-resolution imagery, such as *Landsat* MSS, that was available at the time and in use by the image-processing community. However, we now have access to large-scale, high-resolution imagery that allows for the structural analysis of the road surface. In fact, for practical mapping applications, simply estimating the location of the road centre line is inadequate. For practical mapping applications a detailed analysis must be performed in order to maintain positional accuracy and to extract attributes of the road such as surface material, number of lanes, location of overpasses and cloverleaves.

One can categorize previous work in the area of road finder–followers into one of three major types: correlation trackers, region-based followers and edge linkers. All of these methods use a single local tracking strategy to find roads. Therefore a major problem with previous work is that if the method fails at some point it is very difficult to recover. Further, it is often difficult to automatically recognize when the tracking method has failed because the local nature of these methods assumes that the local maximum, no matter how poor, is its best guess for the position of the road.

Our approach in the ARF (a road follower) system (McKeown & Denlinger 1986) for road tracking is to use multiple cooperative methods for extracting information about road location and structure from complex aerial imagery. This system is a multilevel architecture for image analysis that allows for cooperation among low-level processes and aggregation of information by high-level analysis components. Two low-level methods have been implemented; road-surface texture correlation and road edge following. Each low-level method works independently to establish a model of the centre line of the road, and to extract various road properties such as width, surface material changes and overpasses. Intermediate-level processes monitor the state of the low-level feature-extraction methods and make evaluations concerning the success of each method. As a result of these evaluations one tracking method may be suspended due to apparent failure and restarted from the model generated by other successful trackers. Finally, a high-level module generates a symbolic description of the road in terms of various attributes of the road such as centre line, road width, surface material, overpasses and indication of potential vehicles on the road. This description is available in both map and image coordinate systems. Each tracker makes assumptions and uses road-specific knowledge in its image analysis. The surface tracker makes the assumption that the road exhibits nearly constant width and that the surface-intensity profile on the road changes either very gradually or suddenly. Abrupt changes in the surface-intensity profile may indicate surface material change, occlusions by trees or overpasses, or shadows cast by nearby buildings. The edge tracker makes the assumption that the road edges will be locally straight and antiparallel. Both trackers assume that the road exhibits a slowly changing direction and that its path can be modelled locally by a parabola. Figure 4 gives a description of the overall organization of the cooperative road tracker system.

There are several problems inherent in the use of multiple-tracking algorithms simultaneously. The major problem is one of information fusion, that is, the information generated by the two methods may give different estimates of road direction, width, etc. How can we reconcile these conflicting sources of information? A further complication is that we may not be able to determine which method is 'correct' until we proceed further down the road. One common example occurs when the surface correlator encounters a surface-material change. Therefore, it is important for the system to defer decisions and to re-evaluate decisions in the light of new evidence or situations. Both models are maintained even when we are not tracking

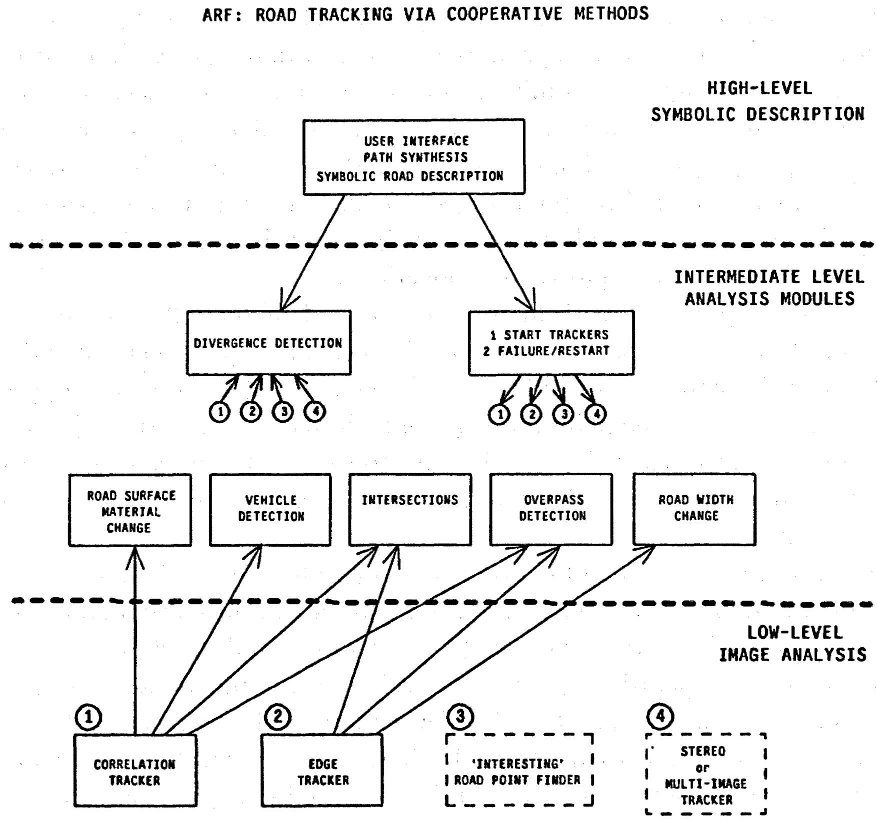


FIGURE 4. ARF system organization.

with one of the methods so that if, for example, the surface tracker does not give good correspondence we can still re-acquire the road. Each model consists of three components: surface cross section, path and edges. Each component is kept consistent with the actual road over the entire road path, whether tracking with both methods simultaneously or tracking with only one. Another major problem is keeping the surface model up-to-date when tracking the edges only. If that is not done properly, it would be possible for the road surface to change while we are tracking only the edges. In this case the road history model will not reflect the actual surface intensity. Figures 5 and 6, plate 2, show two snapshots of the road tracker running on a representative high-resolution aerial image. Note that ARF produces textual descriptions of surface-material changes, overpasses and road-width changes, annotated on the display for the user. The positions of these features are stored in both image and geographical coordinates and can be superimposed by map-to-image correspondence on other imagery of the same road. Figure 5 shows a point where the edge tracker fails, primarily because of surface-material change on a bridge-deck overpass. It is determined by the high-level evaluation module that the two trackers have diverged, that the surface-tracker model is better than the edge tracker, and therefore the edge tracker is restarted from the surface-tracker model. In figure 6 the overpass is detected as well as the bridge-deck surface change, and a width change as the road narrows because of an incoming lane.

In the following section we describe a system, SPAM, that uses spatial and structural constraints represented as production rules to interpret airport and suburban house scenes. SPAM is an example of a high-level knowledge-based interpretation system which uses several





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

image-analysis programs to perform image segmentation and verification, but does not itself directly address image analysis. SPAM acts as a coordinator and evaluator of symbolic descriptions generated by other components performing image analysis.

## 5. Rule-based airport scene interpretation

SPAM, system for photo interpretation of airports by using MAPS, is an image-interpretation system. SPAM (McKeown & McDermott 1983; McKeown *et al.* 1985; McKeown & Harvey 1987) coordinates and controls image segmentation, segmentation analysis and the construction of a scene model. It provides several unique capabilities to bring map knowledge and collateral information to bear during all phases of the interpretation. These capabilities include:

(i) the use of domain-dependent spatial constraints to restrict and refine hypothesis formation during analysis;

(ii) the use of explicit camera models that allow for the projection of map information on to the image;

(iii) the use of image-independent metric models for shape, size, distance, absolute and relative position computation;

(iv) the use of multiple-image cues to verify ambiguous segmentations. Stereo pairs or overlapping image sequences can be used to extract information or to detect missing components of the model.

Figure 7 shows the overall organization of the interpretation system. SPAM maintains an internal spatial database that is composed of features extracted from imagery by various

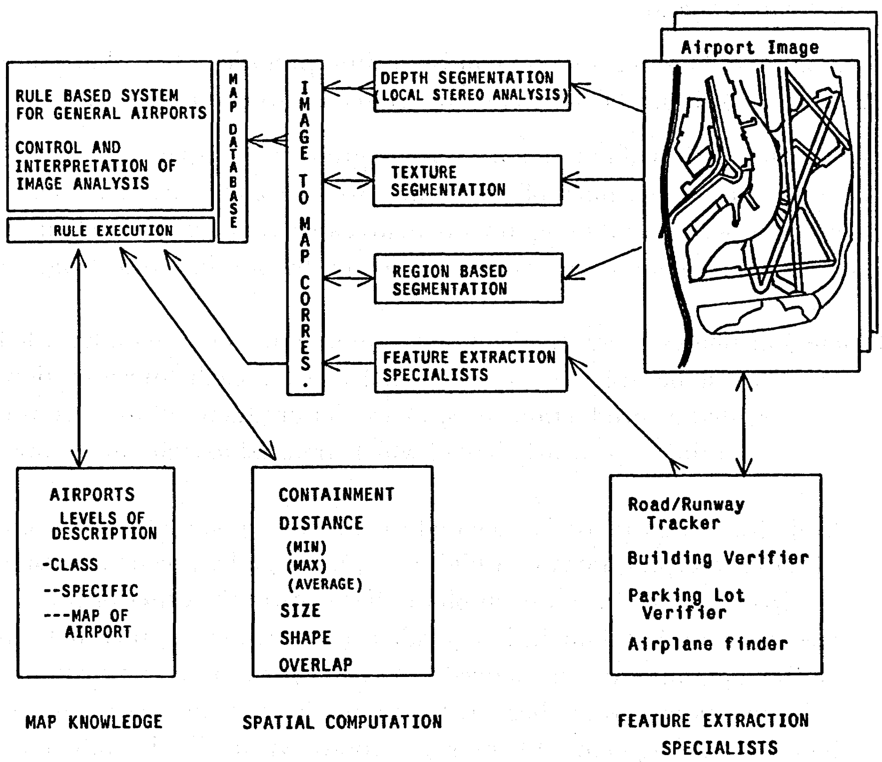


FIGURE 7. SPAM: system organization.

methods, possibly from several images, where the features are represented in terms of their geodetic position rather than their image coordinates. In fact, SPAM performs interpretation in map space which allows for a variety of knowledge such as maps and multitemporal imagery to be handled in a uniform manner. SPAM uses the following components of the MAPS system:

(i) CONCEPTMAP to retrieve shape and position models and site-specific map knowledge;

(ii) map-to-image correspondence to project models on to new imagery and image-to-map correspondence to calculate metric distances and areas;

(iii) procedures to compute spatial relations between hypotheses including containment, intersection, adjacency, closest point of approach, and subsumption.

### 5.1. *The SPAM architecture*

SPAM represents four types of interpretation primitives, regions, fragments, fractional areas and models. SPAM performs scene interpretation by transforming image regions into scene fragment interpretations, aggregating these fragments into consistent and compatible collections called functional areas, and selecting sets of functional areas that form models of the scene. Loosely speaking there are four phases of interpretation.

*Phase* 1: *region-to-fragment*. Assigns the image region data a set of fragment interpretations based solely on local properties (2D shape characteristics, texture, 3D depth, height, etc.) and knowledge about the classes of objects found in the scene.

*Phase* 2: *local-consistency-check*. Pair-wise tests are performed on the fragment interpretations that use spatial knowledge about the scene under consideration. The confidence of those interpretations supporting one another is incremented based on the quality of the test.

*Phase* 3: *functional area*. Sets of mutually consistent interpretations that share similar functions or are spatial decompositions of the scene are grouped into cliques called functional areas.

*Phase* 4: *model-generation*. Sets of functional areas are grouped together into scene segments. The segments with the largest number of functional areas become distinct scene models. Any conflicts encountered when combining functional areas are resolved by a default strategy, using the accumulated support for each interpretation, or by specific knowledge added by the user.

Each of these four phases operationalizes one or more types of domain knowledge. To build a SPAM system we must be able to acquire knowledge for each interpretation phase. This knowledge is represented as production rules. Over 500 productions are currently used in the airport-scene interpretation task, nearly half of which are used to evaluate and propagate local consistency.

As shown in figure 8 each phase is executed in the order given above. SPAM drives from a local, low-level set of interpretations to a high-level, more global, scene interpretation. There is a set of hard-wired productions for each phase that control the order of rule executions, the forking of processes, and other domain-independent tasks. However, this 'bottom-up' organization does not preclude interactions between phases. For example, prediction of a fragment interpretation in functional-area phase will automatically cause SPAM to re-enter local-consistency phase for that fragment. Other forms of top-down activity include stereo verification to disambiguate conflicting hypotheses in model-generation phase and linear alignment in





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

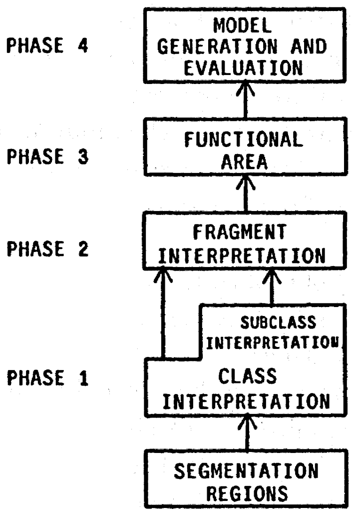


FIGURE 8. Interpretation phases in SPAM.

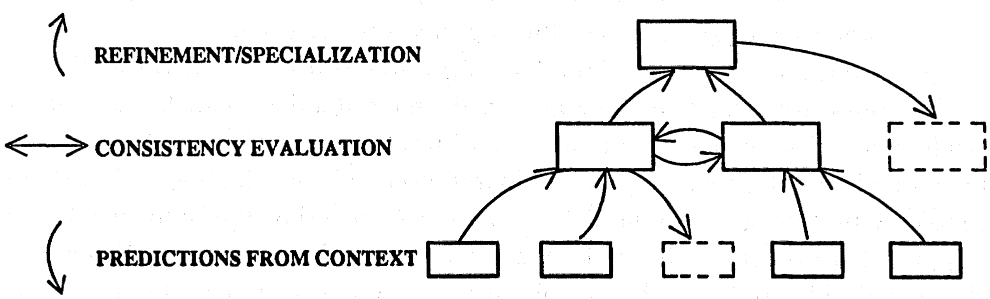


FIGURE 9. Refinement, consistency and prediction in SPAM.

region-to-fragment phase. Figure 9 shows the refinement–consistency–prediction paradigm used in SPAM within each interpretation phase. Knowledge is used to check for consistency among hypotheses, to predict missing components using context, and to create contexts based on collections of consistent hypotheses. Prediction is restrained in SPAM, in that hypotheses cannot predict missing components at their own representation level. A collection of hypotheses must combine to create a context from which a prediction can be made. These contexts are refinements or spatial aggregations in the scene. For example, a collection of mutually consistent runways and taxiways might combine to generate a runway functional area. Rules that encode knowledge that runway functional areas often contain grassy areas or tarmac may predict that certain subareas within that functional area are good candidates for finding such regions. However, an isolated runway or taxiway hypothesis cannot directly make these predictions. In SPAM the context determines the prediction. This serves to decrease the combinatorics of hypothesis generation and to allow the system to focus on those areas with strong support at each level of the interpretation.

### 5.2. *Stereo image analysis*

One image-analysis component of the SPAM system is a stereo-verification module, STEREOSYS (McKeown *et al.* 1986). Stereo verification refers to the verification of hypotheses about a scene by stereo analysis of the scene. Unlike stereo interpretation, stereo verification requires only





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

coarse indications of three-dimensional structure. For aerial photography, this means coarse indications of the heights of objects above their surroundings. STEREOSYS is used within the SPAM system to confirm or refute airport feature hypotheses based on their three-dimensional structure, and operates at a high level of abstraction. The input to STEREOSYS is a segmentation region and two conflicting fragment hypotheses. STEREOSYS returns to SPAM a set of confidence factors that relate the overall confidence of the stereo match and confidence factors for each of the alternative hypotheses.



Stereo verification deals with a variety of problems that are not ordinarily present in isolated experiments with stereo matching and analysis. Some of the most interesting problems within the spatial database context are the following:

(i) an appropriate conjugate image pair has to be selected from a database of overlapping images based on criteria that would maximize the likelihood for good correspondence;

(ii) the image pairs must be dynamically resampled such that the epipolar assumption (i.e. epipolars are scan lines) used in most region-based stereo matching algorithms can be applied;

(iii) an initial coarse registration step is generally necessary because the quality of the correspondence between conjugate pairs varies greatly, in many cases the magnitude of the initial misregistration being greater than the expected disparity shift.

These requirements, in turn, raise a broad set of research issues. In terms of spatial databases the major questions are related to how an aerial image database can be used to generate automatically a useful stereo pair containing an arbitrary region, and how a stereo system can handle the misregistration problems inherent in multisource image databases. The result of this research indicate that image–map database issues in stereo verification influence the utility of such an approach as much as the underlying stereo-matching algorithm. In fact, they are intimately related. The ability to be flexible in the selection of stereo pairs provides opportunities for multitemporal, multiscale or multilook matching. Equally important is flexibility in the matching algorithm, especially with respect to assumptions that require nearly perfectly aligned conjugate images. STEREOSYS uses the following components of the MAPS system.

(i) IMAGE database to select appropriate conjugate-pair imagery bases on time, scale and flightline information.

(ii) Image-to-map correspondence to resample imagery to a common rectified projection so that epipolars align as scan lines in the image.

(iii) TERRAIN database for resampling imagery.



Figure 10, plate 3, shows an area around a hangar at National Airport in Washington, D.C. The left image contains a region of the image under analysis by SPAM. The right image was selected by STEREOSYS as the conjugate pair from the database of images maintained by MAPS. An initial registration was performed to calculate a global disparity shift between the left and right images. This disparity shift corresponds to a displacement of corresponding points in the images due to misregistration between imagery and the MAPS database. The global disparity shift is used to generate a new resampled image that accounts for inaccuracies in the MAPS database and preserves the epipolar matching constraint. The left image and the resampled image are now matched by STEREOSYS using hierarchical correlation and produce a disparity image registered to the original left image. The disparity value is one-to-one correspondence with distance, or depth, from the camera and therefore indicates relative height in vertical aerial photography. In this case, dark pixels indicate points in the disparity image that are close to the observer, light points are further away calibrated towards the ground plane of the scene.

STEREOSYS uses a statistical analysis of the disparity image to generate confidence factors that the region in the left image has low height confidence, 0.051; moderate height confidence, 0.159; and significant height confidence, 0.791. STEREOSYS generates an overall match confidence of 0.818 that confirms the hypothesis of 'hangar building'.

## 6. Conclusions

We have described three examples of image-analysis systems that integrate and use spatial knowledge with image analysis and interpretation techniques. Each example has been shown to rely on a spatial database of geographically referenced facts to provide information crucial to the interpretation task. Each system performs interpretation at a different level of detail, depending on the task and the overall goals of the interpretation system.

I thank the organizers of this conference, particulary Dr B. J. Conway, for a very interesting and enjoyable meeting. I also acknowledge my colleagues at the Digital Mapping Laboratory; particularly J. Denlinger, W. Harvey and J. McDermott for their contribution to the research results reported herein.

## References

Barr, A. & Feigenbaum, E. A. 1981 *The Handbook of artificial intelligence* Los Altos: William Kaufmann.

McKeown, D. M. 1983 MAPS: the organization of a spatial database system using imagery, terrain and map data. In *Proc. DARPA Image Understanding Workshop, June 1983*, pp. 105–127. Arlington, Virginia: Science Applications, Inc.

McKeown, D. M. 1984*a* Digital cartography and photo interpretation from a database viewpoint. In *New applications of Databases* (ed. G. Gargarin & E. Golembe), pp. 19–42. New York: Academic Press.

McKeown, D. M. 1984*b* Knowledge-based aerial photo interpretation. *Photogrammetria* **39**, 91–123.

McKeown, D. M. 1987 The role of artificial intelligence in the integration of remotely sensed data with geographic information systems. *IEEE Trans. Geosci. Remote Sensing* **GE-25** (3), 330–348.

McKeown, D. M. & Denlinger, J. L. 1984 Map-guided feature extraction from aerial imagery. In *Proc. 2nd IEEE Workshop on Computer Vision: Representation and Control. May* 1984. Silver Spring, Maryland: IEEE Computer Society Press.

McKeown, D. M. & Denlinger, J. L. 1986 Cooperative methods for road tracking in aerial imagery. Technical Report CMU-CS-86-175, Computer Science Department, Carnegie-Mellon University, Pittsburgh.

McKeown, D. M. & Harvey, W. A. 1987 Automating knowledge acquisition for aerial image interpretation. Technical Report CMU-CS-87-102, Computer Science Department, Carnegie-Mellon University, Pittsburgh.

McKeown, D. M., Harvey, W. A. & McDermott, J. 1985 Rule based interpretation of aerial imagery. *IEEE Trans. Patt. Analys. Mach. Intell.* **PAMI-7** (5), 570–585.

McKeown, D. M. & McDermott, J. 1983 Toward expert systems for photo interpretation. In *IEEE trends and applications, May* 1983, pp. 33–39. Silver Spring, Maryland: IEEE Computer Society Press.

McKeown, D. M., McVay, C. A. & Lucas, B. D. 1986 Stereo verification in aerial image analysis. *Optical engineering* **25** (3), 333–346.


FIGURE 3. Sequence of image segmentation for small buildings.

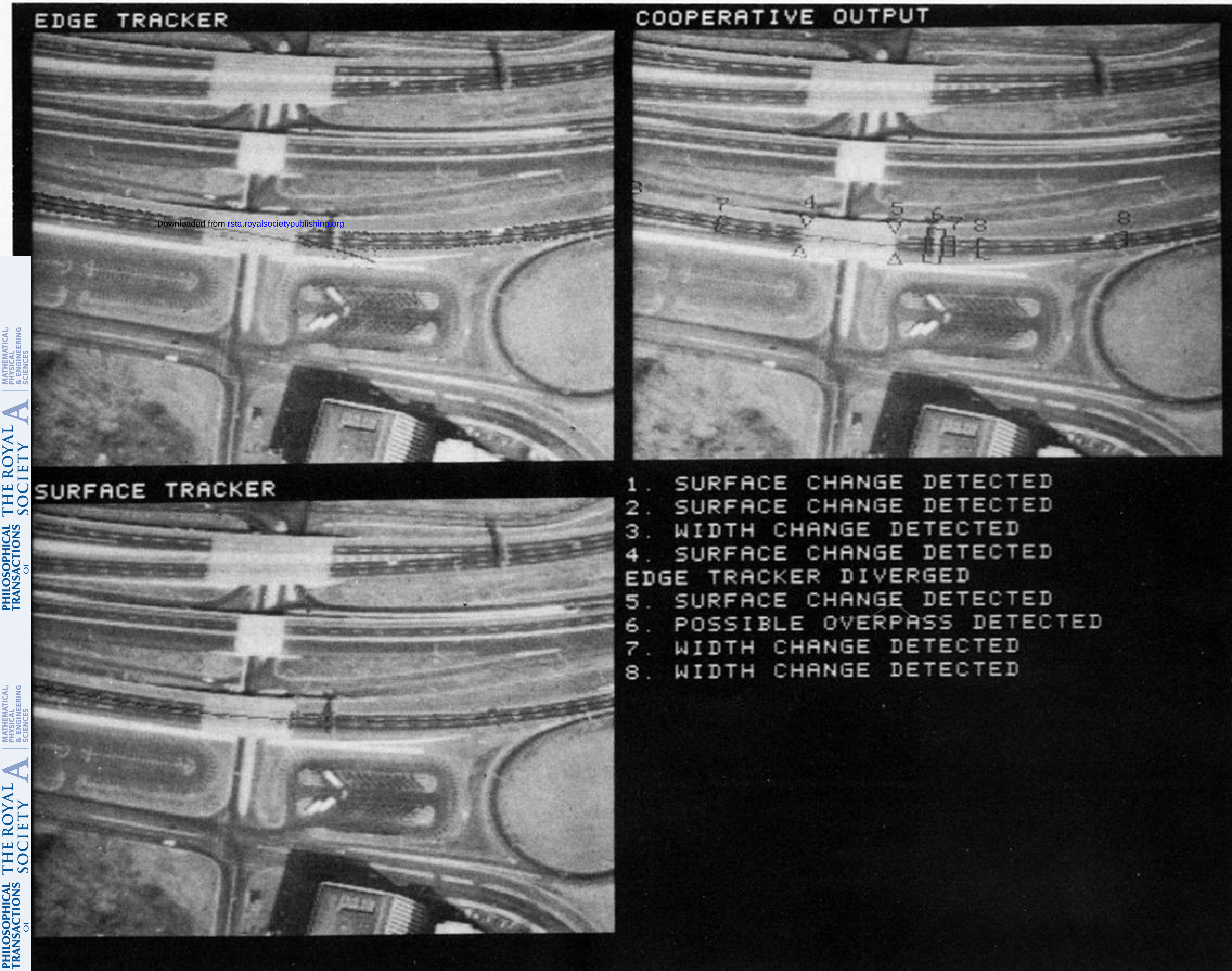


FIGURE 5. ARF: edge tracker fails on low-contrast overpass, successfully restarted from surface tracker.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

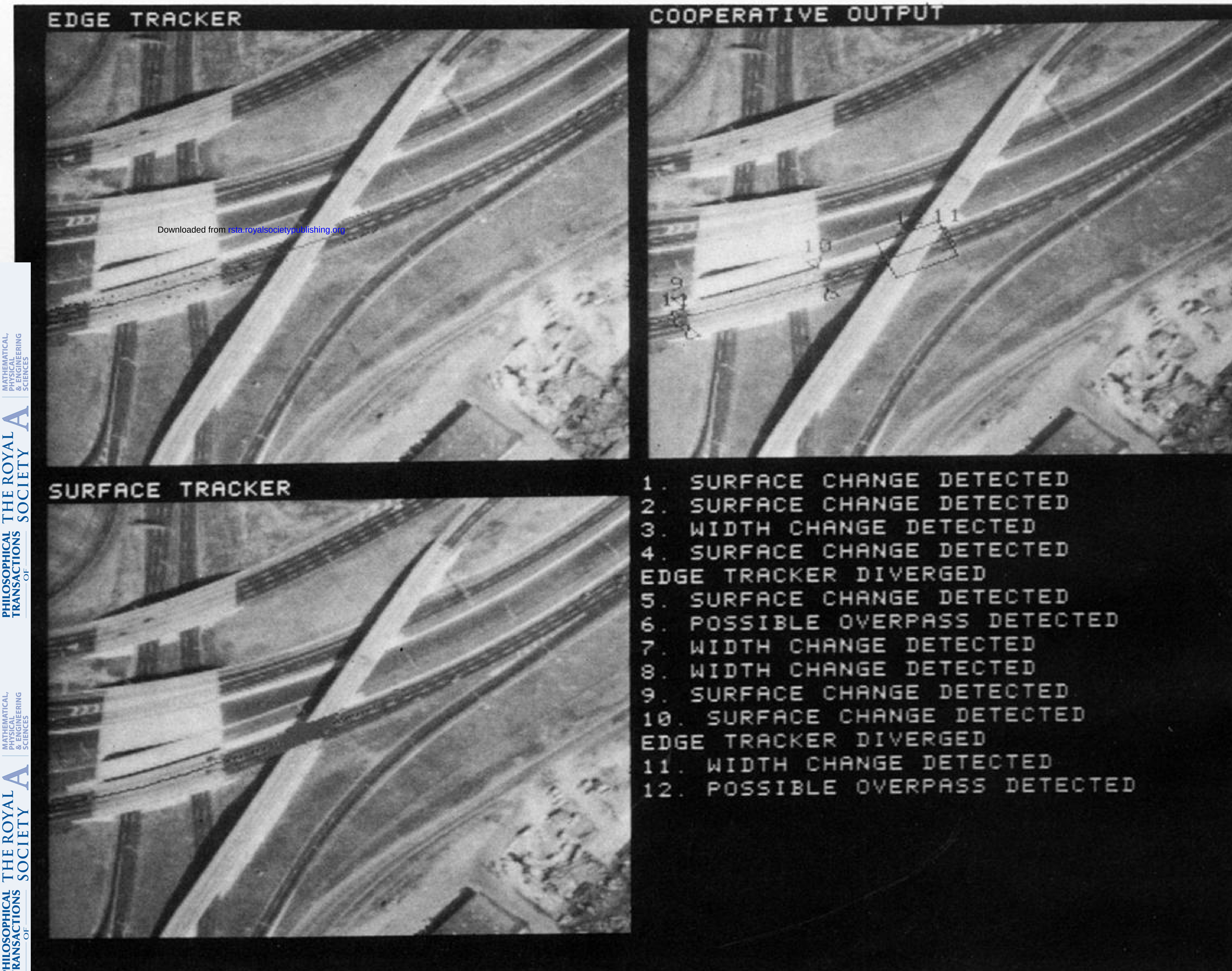


FIGURE 6. ARF: successful detection of overpass and surface material change due to bridge deck.

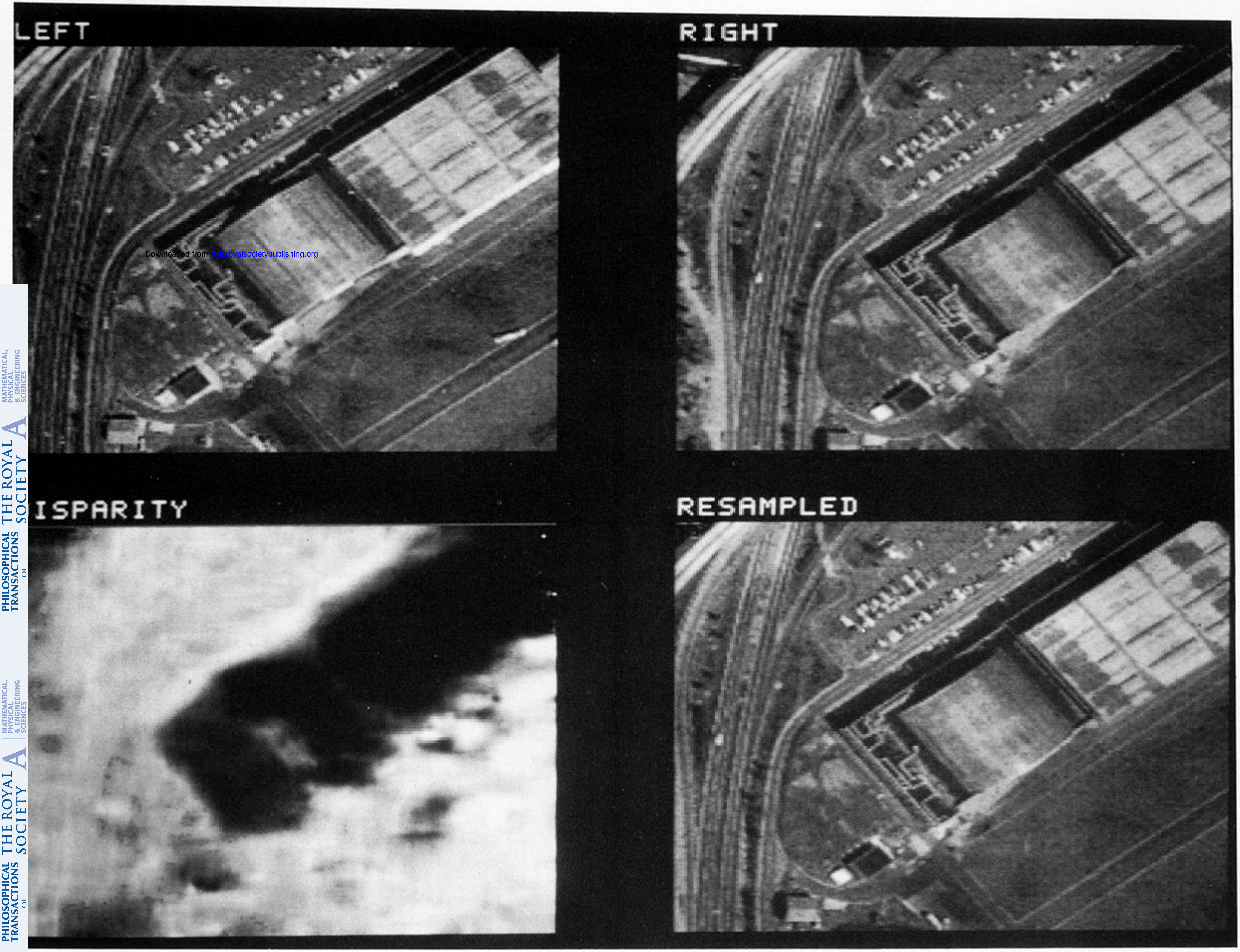


FIGURE 10. Stereo verification in airport scene analysis.